Analogical Information Presentation Method Based on Already Visited Spot for Understanding of Unvisited Area

Michael Shell, *Member, IAENG,* John Doe, *Senior Member, IAENG,* and Jane Doe, *Fellow, IAENG*

Abstract—In recent years, when planning tourist spots, planning is often made by utilizing tourist information on the Web. However, after deciding the area you want to visit from many areas, the user also needs enormous amount of time and effort to find tourist spots that match your image. In addition, there are cases where the user feels expectation and anxiety with respect to the unvisited spot. In this research, in order to support understanding of users' unknown spots, we propose analogy information presentation that supports the understanding by fitting the features of tourist spots that have already visited to unvisited spots. In order to emphasize the features of the tourist spots themselves, extraction of features of each tourist spot is done by work using all reviews of tourist spots entered by the user, all reviews of tourist spots in the target area. We also conduct an evaluation experiment to construct the prototype system and verify the effect of the analogy information between the visited spot and the unvisited spot.

Index Terms—tourist spots, analogy, understanding support, reviews, cosine similarity, TFIDF, harmonic mean.

Fig. 1. Analogical information presentation method based on already visited spot for understanding of unvisited area

I. Introduction

WHEN deciding the travel destination, the traveler selects tourist spots by planning a travel plan, watching tourist spots search sites and books related to tourist information. However, after deciding the area you want to visit from many areas,and further from their many tourist spots in the area is not easy to find. In the case where the tourist spots desired to go are not decided, it is considered that it is more likely to decide tourist spots by looking at ranking and recommendation information. At this time, the image for the tourist spots selected by the user becomes ambiguous, which may cause anxiety.

In recent years, the speed of development of tourism industry and social networking service is accelerating, and the number of users who post reviews on tourist spots experienced to the tourist spot search site is increasing. In order to effectively understand various tourist spots, it is essential to consider the correspondence between unknown information and existing information based on existing information. This way of thinking is equivalent to analogy which applies to the things by previous experiences (called "bases"), or problems (called "targets"). For example, whereas unknown spots such
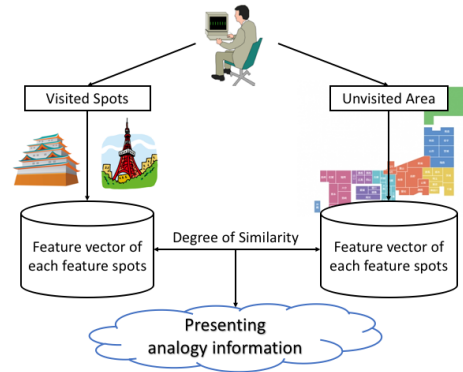
as "Kanazawa's Nisityayagai", if you explain that it is similar to the already visited "Kyoto Hanamikoji", it may make it easier to understand the image.

In this research, in order to support understanding of users' unknown spots, we propose analogy information presentation that supports the understanding by fitting the features of tourist spots that have already visited to unvisited spots. Specifically, from the already visited spot and the unvisited area entered by the user, we use the review to extract the unique features of each spot in the already visited spot and the unique features of each spot in the unvisited area, compare and present analogy information. With this prototype system, users aim to support understanding of unvisited areas. Fig. 1 is a conceptual diagram of the proposed method.

The structure of this paper is as follows. Section II describes related research. Section III gives an overview of the proposed method. Section **??** describes evaluation experiments and considerations to verify the effect of the constructed prototype system. Section V describes with conclusions and future work.

II. Related Work

III. Analogical Information Presentation Method

We propose an analogy information presentation method based on an already visited spot for supporting understanding of unvisited areas. Specifically, first, the user inputs a plurality of tourist spots that have been visited and tourist spot area information that user wishes to visit. Use the already visited spot review vector to find the feature vector for each visited
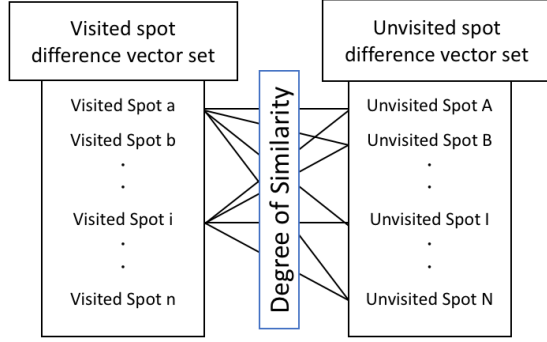
Fig. 2. Concept of similarity calculation

spot. Similarly, the feature vector of each spot in the area is obtained for an unvisited spot. Next, we associate unvisited tourist spots with features similar to the difference features between the visited spot review vector and the unvisited spot review vector. Finally, analogy information for supporting understanding of unvisited areas is defined using TFIDF and presented to the user.

A. Relative features of tourist spots

In this research, features of tourist spots make use of relative features. The relative feature is a unique feature when a specific tourist spot is compared with other tourist spots included in a set of tourist spots. As an example, consider the case where there are "Rokuonji" and "Kiyomizudera" in the tourist spot group. At this time, the features of "Rokuonji" will be gold color, gold leaf, glow, etc., the features of "Kiyomizudera" are the stage, the womb inside, the panoramic view etc. Because both are temples existing in Kyoto, features related to Kyoto and temples do not appear as unique features. Next, consider the case where "the Tokyo Metropolitan Government Building Observatories" and "Rokuonji" exist within the tourist spot group. Features of "Rokuonji" at this time will be Kinkakuji, a temple, golden color, Kyoto etc. Features of "the Tokyo Metropolitan Government Building Observatories" will be perspectives, night view, Shinjuku etc. If the categories of tourist spots are largely different, features as categories will appear. Also, it can show the features of the spot itself. In this research, when a certain spot compares with other spots in the set, we focus on the relative features that make it possible to clarify the features of each spot.

B. Calculation of similarity by cosine similarity

Review vectors of previously visited spots and unvisited spots are created using a discriminated (original) review with the morphological analyzer "mecab-ipadic-NEologd"[1]. After that, using Distributed[2] Bag-of-Words of Doc2Vec, we use a vector created in 300 dimensions using all reviews of each spot. In this paper, we will use the review data obtained from "Jalan"[3] until the end of September 2016.

[1] https://github.com/neologd/mecab-ipadic-neologd/
[2] https://radimrehurek.com/gensim/models/doc2vec.html
[3] https://www.jalan.net/kankou/

The spot differential vector is defined as formula 1. Is the value obtained by subtracting the average value of the spot vectors of the spots of the spot group excluding the spot for which the spot differential vector is found. $spot_{set} = \{s_1, s_2, \ldots, s_i, \ldots, s_n\}$1 is an already visited spot set or an unvisited spot set. s_i is a tourist spot in the set.

$$v_i = s_i - average(spot_{set} - s_i) \qquad (1)$$

From the feature difference vector v_i of the visited spot and the feature difference vector v_j of the unvisited spot, the relative feature similarity between the visited spot and the unvisited spot (Fig. 2). For the similarity calculation, use the cosine scale (formula **??**).

$$cos(v_i, v_j) = \frac{v_{i1}v_{j1} + v_{i2}v_{j2} + \cdots + v_{in}v_{jn}}{\sqrt{v_{i1}^2 + \cdots + v_{in}^2} \times \sqrt{v_{j1}^2 + \cdots + v_{jn}^2}} \qquad (2)$$

A correlation between the already visited spot having similarity of each feature vector of the already visited spot and each feature vector in the unvisited area of 0.125 or more and the highest similarity and the unvisited spot is performed.

C. Feature vector generation by TFIDF

All tourist spots review words by using morphological analyzer "mecab-ipadic-NEologd". However, words obtained by using these words contain words that do not hold Japanese, and it is necessary to delete these noises. Specifically, delete particles, auxiliary verbs, rentaishi, symbols.

Since the analogy information of the visited spot and the unvisited spot associated in section III-B is presented to the user in word format, a review set of a certain tourist spot is assumed to be a document i and a spot where the word j for i appears When the number of occurrences of the set is $TF_{i,j}$, the word j is the number of documents in the spot set is DF_j, and the total number of spots in the spot set is $|D|$ The feature quantity of a word in a spot is defined by the formula 3.

$$word_{i,j} = TF_{i,j} \times IDF_j \qquad (3)$$

$$IDF_j = log(\frac{|D|}{DF_j}) \qquad (4)$$

In this method, for visited spots, the user inputs plurality of spots. By considering all reviews of each spot as one document at a time and by considering all reviews of the other spots as documents, the TFIDF value is calculated by the formula 3, 4, and use it as the feature vector for each spot in the unvisited area.

Regarding the unvisited area, the user designates an area and inputs it. By considering all reviews of each spot in the area as one document and considering all reviews of the other spots as a document, the TFIDF value is calculated by the formula 3, 4, and use it as the feature vector for each spot in the unvisited area.

D. Presenting similar information by harmonic mean

The analogy information for imaging the unvisited spot from the already visited spot is presented to the user in word format. The analogy information of the visited spot and the unvisited spot associated with the section III-C, it is determined using harmonic mean by the feature vector of each spot obtained by the section III-B. The harmonic mean is the reciprocal of the arithmetic mean of the reciprocal. Extracts commonly appearing words in the review document of the already visited spot and the review document of the unvisited spot. The score of the extracted word is defined by the formula 5. $word_{visited}$ and $word_{unvisited}$ indicate the TFIDF value of the visited spot and the TFIDF value of the unvisited spot, respectively. When the value of the word score is large, the TFIDF value of each of the visited spot and the unvisited spot is large, that is, the word has high importance in each document. Therefore, the top ten words of the word score are presented to the user as analogy information.

$$score = \frac{1}{\frac{1}{2}\left(\frac{1}{word_{visited}} + \frac{1}{word_{unvisited}}\right)} \tag{5}$$

IV. Evaluation Experiment

A. Experiment of feature word representing relation

B. Comparative experiment of correspondence

V. Conclusions and Future Work

Acknowledgment

The authors would like to thank...

The preferred spelling of the word "acknowledgment" in American English is without an "e" after the "g." Use the singular heading even if you have many acknowledgments. Avoid expressions such as "One of us (S.B.A.) would like to thank ... ." Instead, write "F. A. Author thanks ... ." Sponsor and financial support acknowledgments are placed in the unnumbered footnote on the first page, not here.

References

[1] 廣嶋 伸章, 安田 宜仁, 藤田 尚樹, 片岡 良治, 地理情報検索におけるクエリ入力支援のための特徴語の提示, 第 26 回人工知能学会全国大会, Vol.26, 1C1-R-5-6, 2012

[2] 松本 敦志, 杉本 徹, クチコミから抽出した特徴語を利用する観光地検索支援, 第 75 回全国大会講演論文集, Vol.2013, No.1, pp.307-308, 2013

[3] 上原 尚, 嶋田 和孝, 遠藤 勉, Web 上に混在する観光情報を活用した観光地推薦システム, 社団法人 電子情報通信学会, 信学技報, Vol.112, No.367, pp.13-18, 2012

[4] 野守 耕爾, 神津 友武, 口コミデータに PLSA を適用した観光客目線による観光地分析, 第 29 回人工知能学会全国大会, Vol.29, 1J2-OS-18a-2, pp.1-4, 2015

[5] Gick, M.L. and Holyoak, K.J.: Analogical Problem Solving, Cognitive Psychology, Vol.12, pp.306–355, 1980

[6] 石田 純太, 砂山 渡, 新知識を理解するための類推能力の育成, 第 30 回人工知能学会全国大会, Vol.30, 3F3-3, 2016

[7] 砂山 渡, 石田 純太, 川本 佳代, 西原 陽子, 類推による説明スキルの獲得支援システム, 情報処理学会論文誌, Vol.59, No.10, 1922–1931, 2018

[8] Jun NAKAMURA, Yukio OHSAWA, Effect of Vacillation on Analogical Thought Process for Concept Creation, Oukan, Vol.2, No.1, p.40-48, 2008

[9] Gentner, D.: Structure-Mapping: A Theoretical Framework for Analogy, Cognitive Science, Vol.7, pp.155–170, 1983

[10] 杉山 将, 確率分布間の距離推定:機械学習分野における最新動向, 日本応用数理学会論文誌,Vol.23, No.3, pp.439-452, 2013